PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04L 12/56, 12/413, G06F 17/00 — A1

(11) International Publication Number: WO 99/40703

(43) International Publication Date: 12 August 1999 (12.08.99)

(21) International Application Number: PCT/AU99/00077

(22) International Filing Date: 5 February 1999 (05.02.99)

(30) Priority Data:
PP 1692 — 6 February 1998 (06.02.98) — AU

(71) Applicants (*for all designated States except US*): ERICSSON AUSTRALIA PTY. LTD. [AU/AU]; 61 Riggall Street, Broadmeadows, VIC 3047 (AU). ROYAL MELBOURNE INSTITUTE OF TECHNOLOGY [AU/AU]; 124 LaTrobe Street, Melbourne, VIC 3000 (AU). CENTRE NATIONAL DE RECHERCHE SCIENTIFIQUE (CNRS) [FR/FR]; 3, rue Michel-Ange, F-75794 Paris Cedex 16 (FR).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): VEITCH, Darryl, Neil [AU/AU]; 44 Erie Avenue, Rowville, VIC 3178 (AU). ROUGHAN, Matthew [AU/AU]; 253 Pigdon Street, North Carlton, VIC 3054 (AU). ABRY, Patrice [FR/FR]; 40, rue de Rhône, F-69007 Lyon (FR).

(74) Agent: CARTER SMITH & BEADLE; 2 Railway Parade, P.O. Box 557, Camberwell, VIC 3124 (AU).

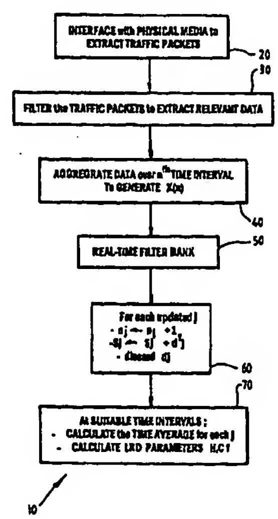
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: REAL-TIME ESTIMATION OF LONG RANGE DEPENDENT PARAMETERS

(57) Abstract

A method and apparatus of estimating long range dependent parameters, such as Hurst parameter H and size parameter c_f, of a data stream in real-time. It includes inputting blocks of data of the data stream to a real-time discrete wavelet decomposition means (200) used to generate wavelet coefficients. A sum of squares of the coefficients at each scale is maintained together with the number of elements combined in the sum within processor means, such as a PC (304). When an estimation is required, averages of the squares of the coefficients are formed, followed by the calculation of a weighted linear regression using the averages leading to a determination of estimates of the parameters.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

REAL-TIME ESTIMATION OF LONG RANGE DEPENDENT PARAMETERS

The present invention relates to a real-time method of estimating parameters used to characterise a Long Range Dependent (LRD) process such as is found in data streams. The present invention more particularly relates to a method of estimating the Hurst (H) parameter and size parameter c_f of a telecommunications data traffic sequence in real-time which assists in the analysis of data in teletraffic applications.

The phenomenon of LRD has recently attracted strong interest in telecommunications with the discovery of self-similar and long range dependent properties in data and communications traffic of diverse types. The investigation of the impact of this on telecommunication network performance has highlighted the need for accurate and computationally effective estimation methods for LRD parameters.

Long range dependence is known to be present in a wide variety of generalised data types including data traffic in high speed telecommunications networks. The presence of LRD allows one to predict trends in data traffic where there is a strong correlation between sets of the data. A common definition of LRD is the slow, power-law like decrease at large lag of the autocovariance function of a stationary stochastic time series $\{x_t\}$. Equivalently, it can be defined as the power-law divergence at the origin of the spectrum of that series:

$$f_x(\upsilon) \sim c_f |\upsilon|^{-\alpha}, |\upsilon| \to 0 \qquad (1)$$

or

$$\sim c_f |\upsilon^{1-2H}, |\upsilon| \to 0$$

where $H = (1 + \alpha)/2$ is the Hurst parameter, $0.5 < H < 1$, α is the "scale" parameter and c_f is the "size" parameter.

Essentially the LRD phenomenon states that the sum of all correlations downstream from any given time instant is always appreciable, even if individually the correlations are small. Crucially, it implies that there is no possibility of defining a characteristic time-scale beyond which correlations would have essentially disappeared, as would be the case for a process whose autocorrelation function decayed exponentially - the classical assumption. Thus, one cannot find a reference unit of time over which, for instance, some property of the data could be reliably measured. Instead of a single prominent time scale, LRD is characterised by scale invariance properties governed by the parameter α which describes the relationship between scales.

The simplest definition of LRD involves the two parameters α and c_f mentioned hereinbefore, of which α is more important.

As α, and therefore H, appear in the exponent of (1) it defines the existence of the LRD phenomenon and governs the characteristic scaling behavior of a LRD process as well as statistics derived from it including basic ones such as the sample mean of the series. Thus H gives a measure of the rate of decay of correlation between sets of data. It is therefore important that α is estimated well.

The parameter c_f plays a major role in fixing the absolute size of LRD generated effects, the general character of which is determined by H. Estimating c_f is an issue of importance for quantitative analysis, but is fraught with the same statistical difficulties intrinsic to H estimation.

There exist estimators of the Hurst parameter that are either time-domain based or frequency-domain based which have poor statistical performance including high bias and/or high variance.

However, estimators using wavelet analysis can provide a natural, statistically and computationally efficient estimation of the Hurst parameter H, in an unbiased

manner. Wavelet analysis is a tool which studies the scale dependent properties of data directly via the coefficients of a joint scale-time wavelet decomposition. The wavelet decomposition takes into account the scaling behavior of a process by examination over a multitude of scales. As such, very little needs to be assumed about the underlying process. Should evidence of LRD be found, it then offers an unbiased semi-parametric estimator which can be efficiently implemented using techniques from discrete multi-resolution analysis (MRA) (see P. Abry, P. Goncalvès and P. Flandrin - "Wavelets, Spectrum Estimation, 1/f Processes", Wavelets and Statistics, Lectures Note in Statistics, Vol. 105 (1995), pp. 15-30).

However, such wavelet estimators only take into account the "static" analysis of a collection of data up to a particular time instant or over a time period. As such, it is an off-line process requiring a vast amount of memory to store each and every data set for the analysis thereof. If new sets of data are required to be added to the existing collection of data and thereafter analysed, all of the data sets are still required to be maintained and therefore more memory is required for the storage of the data sets. As further data sets are added, it becomes impractical to have vast arrays of memory storage devices.

There is a need to be able to analyse real-time data, for example, teletraffic data so that the effect of additional data on a telecommunications network can be considered instantly without the requirement of storing vast amounts of data sets over a particular time interval.

The present invention provides for a real-time method of estimating LRD parameters of on-line data streams based on wavelet analysis without the need for large memory storage and rapidly enough to handle very high data rates.

The method scales naturally so that as data transfer rates become higher over time, the method will be implementable and effective in terms of speed and memory storage.

According to a first aspect of the invention there is provided a method of estimating long range dependent (LRD) parameters, such as Hurst parameter H and size parameter c_f ,of a data steam in real-time, said method comprising the steps of:

inputting each block of data of said data stream to a real-time discrete wavelet decomposition means;

extracting wavelet coefficients generated from said wavelet decomposition means for each block of input data;

maintaining a sum of squares of said wavelet coefficients and maintaining the number of elements combined in said sum at each one of a number of scales;

at times when an estimation is required, the method further comprising the following steps:

forming averages of said squares of said wavelet coefficients at said each one of a number of scales;

performing a weighted linear regression using said averages over a range of scales; and

determining estimates for the LRD parameters on the basis of said linear regression.

According to a second aspect of the invention, there is provided a method of estimating LRD parameters, such as H and c_f , of a data stream in real-time, said method comprising the steps of:

extracting packets of data from said data stream;

aggregating the packets over an n^{th} time interval to generate a data point x (n); inputting each data point x(n) to a real-time discrete wavelet decomposition means;

extracting wavelet coefficients d_j generated from said decomposition means for each data point x(n);

maintaining a sum S_j of squares d^2_j of said wavelet coefficients and maintaining a number n_j of data points used in said sum S_j at each scale $_j$ of a number of scales;

at times when an estimation is required, the method further comprising the following steps:

forming averages of the squares d^2_j of said wavelet coefficients from said sum S_j at each scale $_j$;

performing a linear regression using said averages over a range of scales; and

determining estimates for the LRD parameters on the basis of said linear regression.

The method may further include calculating confidence intervals for said estimates. Preferably, the wavelet coefficients are discarded after the step of squaring said wavelet coefficients. Preferably, prior to the performing step, a range of scales is chosen over which the linear regression is plotted. The step of choosing the range of scales may be performed manually on long time scales or automatically according to a suitable algorithm.

The wavelet decomposition means may comprise a multi-resolution algorithm means or general filter bank means. The wavelet decomposition means may be implemented in software or hardware, for example, using a digital signal processing chip.

The present invention also provides for apparatus for estimating LRD parameters, such as H and c_f, of a data stream in real-time, said apparatus comprising:

means for receiving data packets of said data stream;

pre-processor means for aggregating the received data packets over time intervals to generate respective data points x(n);

a real-time discrete wavelet decomposition means for receiving each data point x (n);

said decomposition means generating wavelet coefficients d_j for each received data point x(n);

means for calculating a sum S_j of squares d^2_j of said coefficients and deriving a number n; of data points used in said sum S_j at each scale j of a number of scales;

memory means for storing S_j and n_j;

wherein at times when an estimation of the LRD parameters is required, said means for calculating (a) forms averages of the squares d^2_j of said coefficients from said sum S_j at each scale j;

(b) derives a linear regression using said averages over a range of scales, and

(c) determines estimates for the LRD parameters on the basis of said linear regression.

A preferred embodiment of the invention will hereinafter be described with reference to the accompanying drawings wherein:

Figure 1 is a block diagram showing the algorithm used to calculate the LRD parameters from a data traffic stream;

Figure 2 is a schematic diagram of a wavelet decomposition means in the form of a pyramidal filter-bank implementing a discrete wavelet transform to determine the wavelet coefficients;

Figure 3 is a block diagram of hardware used to implement the algorithm of Figure 1;

Figure 4 is a flow diagram showing the steps involved in calculating LRD parameters using an off-line algorithm;

Figure 5 is a linear regression plot of y_j versus j, and

Figure 6 is a log-log plot for Ethernet data.

In Figure 1, there is shown a flow diagram 10 depicting a series of steps used for calculating the LRD parameters, H and c_f, in real-time starting from the initial input of the data stream to be analysed.

At step 20, the process commences by interfacing with the physical media carrying ethernet traffic in order to extract the traffic packets. The traffic packets are then filtered at step 30 to extract relevant data. The process then progresses to step 40 where the relevant data is aggregated over the n^{th} time interval to generate the next data point x(n) of the series to be analysed. Each data point x(n) generated is input to a real-time filter bank at step 50 which is used to extract and update the Discrete Wavelet Transform (DWT) coefficients, to be described with reference to Figure 2. After the coefficients have been determined for each data point x(n), then at step 60, each new data point is added to the existing number of data points for each scale j so that (n_j+1) replaces n_j as the new updated total of data points. The sum of squares S_j is also updated to include the new coefficients squared at that scale. Information on the existing coefficient is discarded. Finally at step 70, at suitable time intervals, a calculation of time average μ_j for each j is made and the LRD parameters H and c_f are calculated to be discussed later.

A discrete wavelet transform (DWT) is performed on each data point x(n) wherein a number of wavelet coefficients are produced through a wavelet decomposition process using the real-time filter bank.

The wavelet transform can be understood as a more flexible form of a Fourier Transform, where the original signal is transformed, not into a frequency domain, but into a time-scale wavelet domain. The sinusoidal functions of Fourier theory are replaced by wavelet basis functions generated by simple translations and dilations of the the mother wavelet ψ_0, itself defined via multiresolution theory (see I. Daubechies, "Ten Lectures on Wavelets", SIAM (1992)). The wavelet transform can be thought of as a method of simultaneously observing the signal at a full range of different scales or resolutions j. The wavelet coefficients of each data point x(n) essentially comprise details d_j and approximations a_j for each scale j.

In Figure 2, there is shown a wavelet decomposition means 200 which may be a multi-resolution algorithm in the form of a real-time filter bank. Each data point

$x(n)$ in the series to be analysed over an interval of time is input to the wavelet decomposition means 200 and fed through band-pass filter (BPF) 202 and downsampler 204. The wavelet coefficient, or detail, of the data point is output at the first scale as d_1. Part of the input data point $x(n)$ is fed to low-pass filter (LPF) 206, through downsampler 208 and then is split again. One portion is input to BPF 210 and downsampler 212 to extract the coefficient d_2 at scale 2. Another portion is input to LPF 214 and downsampler 216. The process is repeated to extract the coefficients up to scale j. Therefore, output d_1 is updated for every second new data point $x(n)$, output d_2 is updated for every fourth new $x(n)$ and output d_j is updated for every $(2^j)^{th}$ new $x(n)$.

Each BPF and LPF in the wavelet decomposition means 200 is of a finite length K, so that only K input values are held in the memory of each BPF or LPF. Once a value, for example $x(n)$ itself which is input to BPF 202, has propagated through the filter, the value is dropped or discarded. Storage requirements are therefore fixed for each filter with $K. \log_2(n)$ values being stored in the filter bank overall. Each of the filters may be implemented by a Finite Impulse Response (FIR) filter. The wavelet decomposition means may be implemented in software or in hardware, for example, on a DSP chip.

Referring to Figure 1, after the DWT coefficients d_j are updated, using $x(n)$, in the real-time filter bank in step 50, a number of actions occur in step 60. For each $x(n)$ that is updated at each scale j:

- the number of wavelet coefficients at that scale is incremented by one, that is, n_j is replaced by (n_j+1);
- the existing sum S_j of squared coefficients is replaced by $(S_j + d_j^2)$ to produce an updated sum of squares. The coefficient value d_j is then discarded as it is no longer required.

Therefore, all that is required to be stored in memory is the updated sum S_j and the number of data points n_j. Note that it is only when a d_j is updated in step 50, that it

is forwarded onto the next step 60 for the calculation of the new n_j and S_j and then d_j is discarded thereafter. The new values n_j, S_j are stored in memory residing in the RAM.

In step 70, at various time intervals suitable to the user, a time average μ_j is calculated for each j, where

$$\mu_j = \frac{1}{n_j}\left(\Sigma\, d_j^{\,2}\right) = \frac{S_j}{n_j} \qquad (2)$$

using the stored S_j and n_j for each scale. Then, based on the values for (n_j) and (μ_j), an estimation of the LRD parameters H and c_f is performed using a wavelet-based joint estimator or off-line algorithm, to be described hereinafter.

Figure 3 shows the hardware, in block form, used to implement the algorithm of Figure 1. A Network Interface Card (NIC) 302 is used to interface with, and capture traffic packets from the Ethernet 300. The Ethernet may be 10 Base T or 10 Base 2. It is then presented to PC 304 which comprises an Intel Pentium PC running a version of the UNIX(R) operating system named Free BSD. The PC 304 runs the traffic analysis software, written in C and performs the necessary calculations for obtaining a sum S_j of squares d^2_j, deriving the n_j of data points used in the sum S_j. Such software may also be used to form averages of d^2_j at each scale j, subsequently derive a linear regression, determine an estimate for H and c_f and calculate any updates for S_j and n_j on arrival of new data points x(n) Alternatively, hardware such as DSP chips may be used to calculate a sum S_j of squares d^2_j, derive n_j and form averages μ_j of d^2_j at each scale j.. The PC 304 includes a software preprocessor 306, a DWT and Estimator Unit 308 and memory unit 310. The preprocessor 306 preprocesses the traffic measurements through the use of the Berkeley Packet Filters, and outputs a time series of the number of bytes per time interval. The preprocessed traffic measurements are passed to unit 308 which

updates the on-going wavelet decomposition, summary statistics and Hurst parameter estimation, and outputs the results periodically to a printer, plotter or other device. The memory unit 310 is in the form of RAM and is used to store the preprocessed traffic measurements, wavelet coefficients derived from the measurements and summary statistics. Note that only the N most recent measurements and coefficients at each scale are required, where N is the length of the FIR filters used to implement the LPFs and BPFs used in the wavelet decomposition. The remainder of the data has been condensed and stored in the summary statistics.

In Figure 4, there is shown a flow diagram 400 of the steps involved in estimating LRD parameters H and c_f using the wavelet-based joint estimator described in "A wavelet-based Joint Estimator of the Parameters of Long-Range Dependence, Technical Report SERC-0043, 1997 by D. Veitch and P. Abry". The steps 410 to 470 are self-explanatory and what follows is a detailed description of the steps.

<u>The Wavelet Estimator</u>

In the analysis of the LRD phenomenon, the following two features, (F1,F2), of the family of wavelet basis functions play key roles:

F1: The family of wavelet basis functions generated from the mother wavelet ψ_0 are constructed from the dilation or change of scale operator:

$$\psi_{j,0}(t) = 2^{-j/2}\,\psi_0(2^{-j}t) \qquad (3)$$

This means that the analyzing family exhibits, by construction, a scale invariance feature. The LRD phenomenon can be understood as the absence of any characteristic frequency (and therefore scale) in the range of frequencies close to the origin. The LRD property can thus be interpreted as a scale invariance characteristic which is efficiently analysed by wavelets.

F2: ψ_0 has a number N of zero or vanishing moments which can be freely chosen provided $N \geq 1$. By definition this means that $\int t^k \psi_0(t)\, dt \equiv 0$, $k = 0, \ldots, N-1$ (but not for $k \geq N$), or equivalently that the Fourier Transform of ψ_0 satisfies $|\Psi_0(\upsilon)| = O(|\upsilon|^N)$ at the origin. This property can be used to control divergences arising with processes having power-law spectra at the origin.

For a process with a power-law spectrum such as a LRD process, these features engender the following key properties of the wavelet coefficients d_j over a range of scales 2^j, $j = j_1 \ldots j_2$ where the power-law scaling holds.

P1: Due to F1, the scale invariance (the power-law behavior) is captured exactly:

$$\mathbb{E} d_j^2 = 2^{j\alpha} c_f C \tag{4}$$

where

$$C = \int |\upsilon|^{-\alpha} |\Psi_0(\upsilon)|^2 d\upsilon. \tag{5}$$

This exact recovery of a power-law is not a trivial effect and results directly from the dilation operator underlying the design of the wavelet basis. Time-frequency or periodogram based estimates would not exhibit such a feature.

P2: Due to F1 and F2, the d_j are a collection of random variables which are quasi-decorrelated (see "Wavelet Analysis and Synthesis of Fractional Brownian Motion" by P. Flandrin, IEEE Trans. on Info. Theory IT-38 (1992) pp. 910-917). In particular, the long-range dependence present in the time domain representation is completely absent in the wavelet coefficient plane $\{j,k\}$.

Elaborating on Property P2 it has been shown that correlations in the time-scale plane decay at least hyperbolically in all directions with exponents controlled by the number of vanishing moments and corresponding to short range dependence. Since by definition the octave $j = \log_2$ (scale), this implies exponential decay in octave j.

From hereon $\log_2$ will denote the logarithm base 2, whereas ln will denote natural logarithms.

The intuitive basis of the estimator can be found by analysing equation (4). Rewriting it as $\log_2 (\mathbb{E}d_x(j,.)^2) = j\alpha + \log_2(c_f C)$ strongly suggests a linear regression approach for estimating (α, c_f), where clearly the slope of the regression would estimate α and the intercept would be related to c_f. The idea of using a log-log plot is common to many contexts when an exponent is the object of interest. The real issue is to what extent the promise of this simple linear form is realised in the resulting estimator, once the inevitable complications are taken into account.

The first, essential, complication is of course that $\mathbb{E}d_j^2$, a second order quantity that can be related to the spectrum of x, is not known but must be estimated. In the present context this is the principal difficulty as it is well known (see "Statistics for Long-Memory Processes", J. Beran, Chapman & Hall (1994)) that the estimation of second order (and other) quantities in a long range dependent context is a delicate task. Here, however, property P2, the quasi-decorrelation of the d_j allows us to effectively use the simple "time average"

$$\mu_j = \frac{1}{n_j} \Sigma\, d_j^2 \qquad (6)$$

where n_j is the number of coefficients at octave j available to be analysed. This quantity is an unbiased and consistent estimator of $\mathbb{E}d_j^2$. (Note that μ_j is the

sample variance of d_j, since from F2 the expectation of d_j is identically zero for each j).

Thus the sample data d_j for n_j samples at scale j of the wavelet decomposition is squared and then summed and divided by n_j to form the "time average".

The second complication is the non-linearity introduced by the $\log_2$, which biases the estimation. We will see below how this problem also can be circumvented under reasonable hypotheses. Simplifying things slightly, we confirm that the fundamental approach underlying our estimator is indeed a linear regression of $\log_2(\mu_j)$ on $\log_2(2^j) = j$. A weighted linear regression will be used as the variances of the $\log_2(\mu_j)$ vary with j.

Linear regression

The fundamental hypothesis of linear regression is $\mathbb{E}y_j = bx_j + a$. We define the quantities $S = \Sigma\ 1/\sigma^2_j$, $S_x = \Sigma\ x_j/\sigma_j^2$ and $S_{xx} = \Sigma x_j^2/\sigma_j^2$ where σ^2_j is an arbitrary weight associated with y_j. The usual unbiased estimator $(\hat{b},\hat{a})$ of (b, a) is

$$\hat{b} = \frac{\Sigma y_j (S x_j - S_x)/\sigma_j^2}{SS_{xx} - S_x^2} \equiv \Sigma w_j y_j, \qquad (8)$$

$$\hat{a} = \frac{\Sigma y_j (S_{xx} - S_x x_j)/\sigma_j^2}{SS_{xx} - S_x^2} \equiv \Sigma v_j y_j, \qquad (9)$$

where the weights w_j and υ_j satisfy $\Sigma w_j = \Sigma j\upsilon_j = 0$, $\Sigma j w_j = \Sigma \upsilon_j = 1$. Note that these conditions imply that there are always both positive and negative υ_j and w_j.

If in addition the y_j are mutually independent then the covariance matrix is given by

$$Var(\hat{b}) = \sum \sigma_j^2 w_j^2 = \frac{S}{SS_{xx} - S_x^2}, \qquad (10)$$

$$Var(\hat{a}) = \sum \sigma_j^2 v_j^2 = \frac{S_{xx}}{SS_{xx} - S_x^2}, \qquad (11)$$

$$Cov(\hat{a}, \hat{b}) = \sum \sigma_j^2 w_j v_j = \frac{-S_x}{SS_{xx} - S_x^2}, \qquad (12)$$

$$r = -S_x / \sqrt{SS_{xx}}, \qquad (13)$$

where r is the correlation coefficient. If $x_j \geq 0$ for each j it is easy to see that r will be negative, and large in magnitude if x_1 is large, as a small change in the slope "to the right" will result in an amplified change of opposite sign in the intercept.

Finally, if we set $\sigma_j^2 = Var(y_j)$, then $(\hat{b}, \hat{a})$ is the minimum variance unbiased estimator (MVUE) (see "Fundamentals of Statistical Signal Processing" by S.M. Kay, Prentice-Hall (1993)) with covariance matrix as above.

Note that in the event of small errors in the values of the σ_j^2 and small correlations between the y_j, the estimator remains unbiased and its covariance matrix can be accurately estimated by the expressions just given.

Thus far we have indicated that $\log_2(\mu_j)$ is the variable y_j of the desired linear regression satisfying $\mathbb{E}y_j = bj + a$. Since $\mathbb{E}\log_2(\mu_j) \neq \log_2(\mathbb{E}\mu_j) = j\alpha + \log_2(c_f C)$ in general, this cannot be exactly true, although under the conditions H1-H3 below, and also assuming n_j large, it can be established that

$$\log_2(\mu_j) \overset{d}{=} N\left(j\alpha + \log_2 c_f C, \frac{2^{j-1}}{n \ln^2 2}\right)$$

where $\overset{d}{=}$ signifies equality in distribution and $N(\mu,\sigma^2)$ is a Gaussian random variable. In a LRD context however, the large scales are usually the most important to consider, and it is precisely there that the n_j are not large. Here the condition$_j$ is removed by examining the distribution of $\log_2(\mu_j)$ in more detail. It turns out that this refinement leads to only a small improvement in the estimation of α, but has very important implications for the estimation of c_f.

Throughout the analysis it is instructive to bear in mind that the number of available detail coefficients n_j essentially decreases by half as the scale is doubled, that is $n_{j+1} \approx n_j/2$, and therefore that $n_{j+1} \approx n2^{-j}$ where n is the length of the initial data.

We assume that the following supplementary hypotheses hold true.

H1: The process x, and hence the processes d(j,.), are Gaussian.

H2: For fixed j the process d(j,.) is iid.

H3: The processes d(j,.) and d(j',.), $j \neq j'$, are independent.

Hypothesis H1 is justified by numerical evidence which shows that the method is very insensitive to the form of the marginal distributions of x. Hypotheses H2 and H3 are both well justified by property P2 (they are separated to make it clearer which properties are needed where).

These extra conditions, whilst appearing very restrictive at first glance, are in fact very reasonable in practical terms, as borne out in simulations. The reason for this is that the underlying effectiveness of the method is based on P1 and P2, H1-H3 being added only to extend the quantitative analysis.

Let the density of a Chi-squared variate $X_\upsilon \overset{d}{=} X_\upsilon^2$ be denoted by

$f_\nu(x) = \left(\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\right)x^{\nu/2-1}e^{-x/2}$. The mean and variance of such a variate are υ and 2υ respectively. Also set $z_j = 2^{j\alpha} c_f C$. From H1 and H2 and equations 4 and 6 we have

$$\mu_j \overset{d}{\sim} \frac{z_j}{n_j} X_{n_j}, \qquad (14)$$

where $\mathbb{E}\mu_j = z_j$ as μ_j is unbiased, and therefore

$$\log_2(\mu_j) \overset{d}{\sim} \log_2 z_j - \log_2 n_j + \log_2 X_{n_j}$$

$$\overset{d}{\sim} j\alpha + \log_2 c_f C - \log_2 n_j + \ln X_{n_j} / \ln 2. \qquad (15)$$

Thus the study of $\log_2(\mu_j)$ reduces to that of the logarithm of a Chi-squared variable.

Using the relations $\int_0^\infty x^{\upsilon-1} e^{-\mu x} \ln x \, dx = \frac{1}{\mu^\upsilon}\Gamma(\upsilon)[\psi(\upsilon) - \ln\mu]$, $\mathrm{Re}\mu > 0, \mathrm{Re}\upsilon > 0$

(see Table of Integrals, Series and Products", I.S. Gradshteyn and I.M. Ryzhik, Academic Press, corrected and enlarged edition (1980)), equation ∮ 4.352), and $\int_0^\infty x^{\nu-1} e^{-\mu x} (\ln x)^2 \, dx = \frac{1}{\mu^\nu}\Gamma(\nu)[(\psi(\nu) - \ln\mu)^2 + \zeta(2,\nu)]$, $\mathrm{Re}\mu > \mathrm{Re}\upsilon > 0$, (above reference equation 2, ∮ 4.358) where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the Psi function and $\zeta(z,\upsilon)$ is a generalised Riemann Zeta function, it is straightforward to show from the definition of $f_\upsilon(x)$ above that

$$IE \ln X_\upsilon = \psi(\upsilon/2) + \ln 2, \qquad (16)$$

$$Var(\ln X_\upsilon) = \zeta(2, \upsilon/2). \qquad (17)$$

It follows that

$$IE \log_2(\mu_j) = j\alpha + \log_2 c_f C + g_j, \qquad (18)$$

$$Var(\log_2(\mu_j)) = \zeta(2, n_j/2)/\ln^2 2, \qquad (19)$$

where the term

$$g_j = \psi(n_j/2)/\ln 2 - \log_2(n_j/2), \qquad (20)$$

a negative function of n_j only, can be easily calculated for all values of n_j.

The term g_j is a small corrective term which is substracted to account for the distorting non-linearity introduced by the log.

For future reference, below are the asymptotic form for n_j large of the quantities above:

$$g_j \sim \frac{-1}{n_j \ln 2} \qquad (21)$$

$$Var(\log_2(\mu_j)) \sim \frac{2}{n_j \ln^2 2} \qquad (22)$$

Defining the variable y_j as $y_j \equiv \log_2(\mu_j) - g_j$ (23)

we see that from the above description it is clear that under H1 and H2 they obey

$$IEy_j = j\alpha + \log_2 c_f C, \qquad (24)$$

$$Var(y_j) = \zeta(2, n_j/2)/\ln^2 2, \qquad (25)$$

and thus satisfy the requirements for a weighted linear regression.

A weighted regression estimation $(\hat{b}, \hat{a})$ of y_j on $j = x_j$ is performed according to equations 8 and 9 with $\sigma_j^2 = Var(y_j)$.

An example of the regression fit using a simulated data set is given in Figure 5 where $y_j = \log_2(\mu_j)$-g_j is plotted against j and showing 95% confidence intervals. The 95% confidence intervals for each y_j, shown as vertical lines at each octave j, are seen to increase with j. This can be understood from equation (22), remembering that $n_j \approx n2^{-j}$ meaning that the number of data points halves with every increase in j by one. The intervals are derived from the known sample variances σ_j^2 of the estimates y_j under gaussian assumptions. An LRD process is apparent between scales 4 and 10 whereas strong SRD or short range dependence is apparent for scales less than 4, and particularly for the range j=1 to j=3. The vertical bars at each octave give 95% confidence intervals for the y_j. The series is simulated farima (0,d,2) with d=0.25 (α=0.50) and $\Psi = [-2, -1]$ implying $c_f = 6.38$. Selecting $(j_1, j_2) = (4,10)$ identifies the relevant scaling range allowing an accurate estimation despite the strong SRD: $\hat{\alpha} = 0.53 \pm 0.7$, $\hat{c}_f = 6.0$ with $4.5 < \hat{c}_f < 7.8$.

In Figure 6, there is shown a log-log plot for real Ethernet data. The data is from an Ethernet trace and contains in excess of 30 million observations which took

only a few seconds to analyse. It plots $\log_2(\mu_j)$ versus j. It shoes an example of the wavelet based scale analysis for continuous time Ethernet data. The asymptotic LRD behaviour is seen to enter at octave j=14. The estimate of the self-similarity parameter $H=1 - \beta | 2$ is H=0.8.

By using the wavelet decomposition means, the static LRD estimation method provides significant advantages in terms of memory storage, memory usage and calculation times. Using it, the input data can be split into blocks, analysed and recombined, so that memory problems are not encountered in treating data of arbitrary length n. The run time complexity of the method is very low, of the order n or O(n), making it very suitable for the analysis of very large data sets. These advantages make the method suitable for real-time implementation where the data is collected and the LRD parameters estimated on a continuous, on-the-fly basis. In the future, when the total amount of data to be processed rises from megabits to gigabits, and even terabits, storage problems will not arise as the requirements of the real-time algorithm vary as the logarithm of the data length.

The real-time version described here and the working preferred implementation prove that these potential real-time advantages can be achieved in practice.

CLAIMS

1. A method of estimating long range dependent (LRD) parameters, such as Hurst parameter H and size parameter c_f ,of a data steam in real-time, said method comprising the steps of:

inputting each block of data of said data stream to a real-time discrete wavelet decomposition means;

extracting wavelet coefficients generated from said wavelet decomposition means for each block of input data;

maintaining a sum of squares of said wavelet coefficients and maintaining the number of elements combined in said sum at each one of a number of scales;

at times when an estimation is required, the method further comprising the following steps:

forming averages of said squares of said wavelet coefficients at said each one of a number of scales;

performing a weighted linear regression using said averages over a range of scales; and

determining estimates for the LRD parameters on the basis of said linear regression.

2. A method according to claim 1 wherein, for each updated scale, the existing sum of squares of said wavelet coefficients is replaced by a new sum of squares equivalent to the existing sum added to the square of each new wavelet coefficient.

3. A method according to claim 1 or claim 2 wherein, for each updated scale, the existing number of elements combined in said existing sum is incremented by the number of new elements at each scale.

4. A method according to claim 2 or claim 3 wherein the value of each coefficient is not retained or stored after each update.

5. A method according to any one of the previous claims wherein each block of data input to the decomposition means is not retained or stored.

6. A method according to any one of the previous claims wherein the step of forming averages includes forming a time average at each scale equivalent to the

updated sum of squared coefficients divided by the updated number of elements combined in the updated sum.

7. A method according to any one of the previous claims wherein the linear regression is plotted and prior to the performing step, a range of scales is chosen.

8. A method according to any one of the previous claims further comprising the step of calculating confidence intervals for each of the derived estimates.

9. A method of estimating LRD parameters, such as H and c_f, of a data stream in real-time, said method comprising the steps of:

extracting packets of data from said data stream;

aggregating the packets over an n^{th} time interval to generate a data point x (n); inputting each data point x(n) to a real-time discrete wavelet decomposition means;

extracting wavelet coefficients d_j generated from said decomposition means for each data point x(n);

maintaining a sum S_j of squares d^2_j of said wavelet coefficients and maintaining a number n_j of data points used in said sum S_j at each scale j of a number of scales;

at times when an estimation is required, the method further comprising the following steps:

forming averages of the squares d^2_j of said wavelet coefficients from said sum S_j at each scale j;

performing a linear regression using said averages over a range of scales; and

determining estimates for the LRD parameters on the basis of said linear regression.

10. A method according to claim 9 wherein, for each updated scale j, the sum S_j is replaced by a new sum $(S_j+d^2_j)$ to yield an updated sum of squares.

11. A method according to claim 9 or claim 10 wherein, for each x(n) updated at scale j, the number of data points n_j, or equivalently the number of coefficients at scale j, is incremented by one such that n_j is replaced by (n_j+1).

12. A method according to any one of claims 9 to 11 wherein the value of each d_j is not retained or stored after each update.

13. A method according to any one of claims 9 to 12 wherein each data point x (n) input to the decomposition means is not retained or stored.

14. A method according to any one of claims 9 to 13 wherein the step of forming averages comprises forming a time average μ_j at each scale j such that

$$\mu_j = \frac{1}{n_j}\left(\Sigma\, d_j^{\,2}\right) = \frac{S_j}{n_j}$$

using the maintained S_j and n_j for each scale.

15. A method according to claim 14 further comprising the step of calculating a random variable y_j where

$$y_j = \log_2(\mu_j) - g_j$$

where g_j is a small corrective term.

16. A method according to claim 15 further comprising the step of plotting y_j against j with confidence intervals about $_j$ based on σ_j where σ_j^2 is an arbitrary weight associated with $_j$.

17. A method according to claim 16 wherein after the plotting step a scaling range is chosen on which to base the linear regression.

18. A method according to claim 17 wherein following selection of the scaling range, a weighted linear regression of y_j on j is performed in the selected scaling range with weight σ_j^2.

19 A method according to claim 18 wherein an estimate of H (or α where α = 2H-1) is obtained from the slope of the regression and an estimate of c_f is obtained from the intercept of the regression.

20. A method according to claim 19 further comprising the step of calculating confidence intervals of the estimate of LRD parameters H and c_f.

21. Apparatus for estimating LRD parameters, such as H and c_f, of a data stream in real-time, said apparatus comprising:

means for receiving data packets of said data stream;

pre-processor means for aggregating the received data packets over time intervals to generate respective data points x(n);

a real-time discrete wavelet decomposition means for receiving each data point x (n);

said decomposition means generating wavelet coefficients d_j for each received data point x(n);

means for calculating a sum S_j of squares $d^2{}_j$ of said coefficients and deriving a number n_j of data points used in said sum S_j at each scale j of a number of scales;

memory means for storing S_j and n_j;

wherein at times when an estimation of the LRD parameters is required, said means for calculating (a) forms averages of the squares $d^2{}_j$ of said coefficients from said sum S_j at each scale j;

(b) derives a linear regression using said averages over a range of scales, and

(c) determines estimates for the LRD parameters on the basis of said linear regression.

22. Apparatus according to claim 21 wherein, for each updated j, the sum S_j , is replaced by a new sum $(S_j + d_j^2)$, to yield an updated sum of squares, which new sum is subsequently stored in said memory means.

23. Apparatus according to claim 21 or claim 22 wherein, for each x(n) updated at scale j, the number of data points n_j , or equivalently the number of coefficients at scale j, is incremented by one , such that n_j is replaced by $(n_j +1)$ and the updated value $(n_j+ 1)$ is subsequently stored in said memory means.

24. Apparatus according to any one of claims 21 to 23 wherein said means for calculating is implemented in software.

25. Apparatus according to any one of claims 21 to 23 wherein said means for calculating calculates the sum S_j, derives n_j and forms said averages using hardware and derives said linear regression and said estimates using software.

26. Apparatus according to any one of claims 21 to 25 wherein the value of each d_j is not retained or stored in said memory means.

27. Apparatus according to any one of claims 21 to 26 wherein each data point x(n) received by said decomposition means is not subsequently retained or stored by said memory means.

28. Apparatus according to any one of claims 21 to 27 wherein said pre-processing means, said decomposition means and said memory means are part of a computing processor such as a PC.

29. Apparatus according to any one of claims 21 to 28 wherein said decomposition means comprises a multi-resolution algorithm means.

30. Apparatus according to claim 29 wherein said multi-resolution algorithm means is a real-time filter bank comprising a series of low-pass filters (LPFs) and band-pass filters (BPFs).

31. Apparatus according to claim 30 wherein said LPFs and said BPFs are finite impulse response filters.

32. A method or system substantially as hereinbefore described with reference to the accompanying drawings.

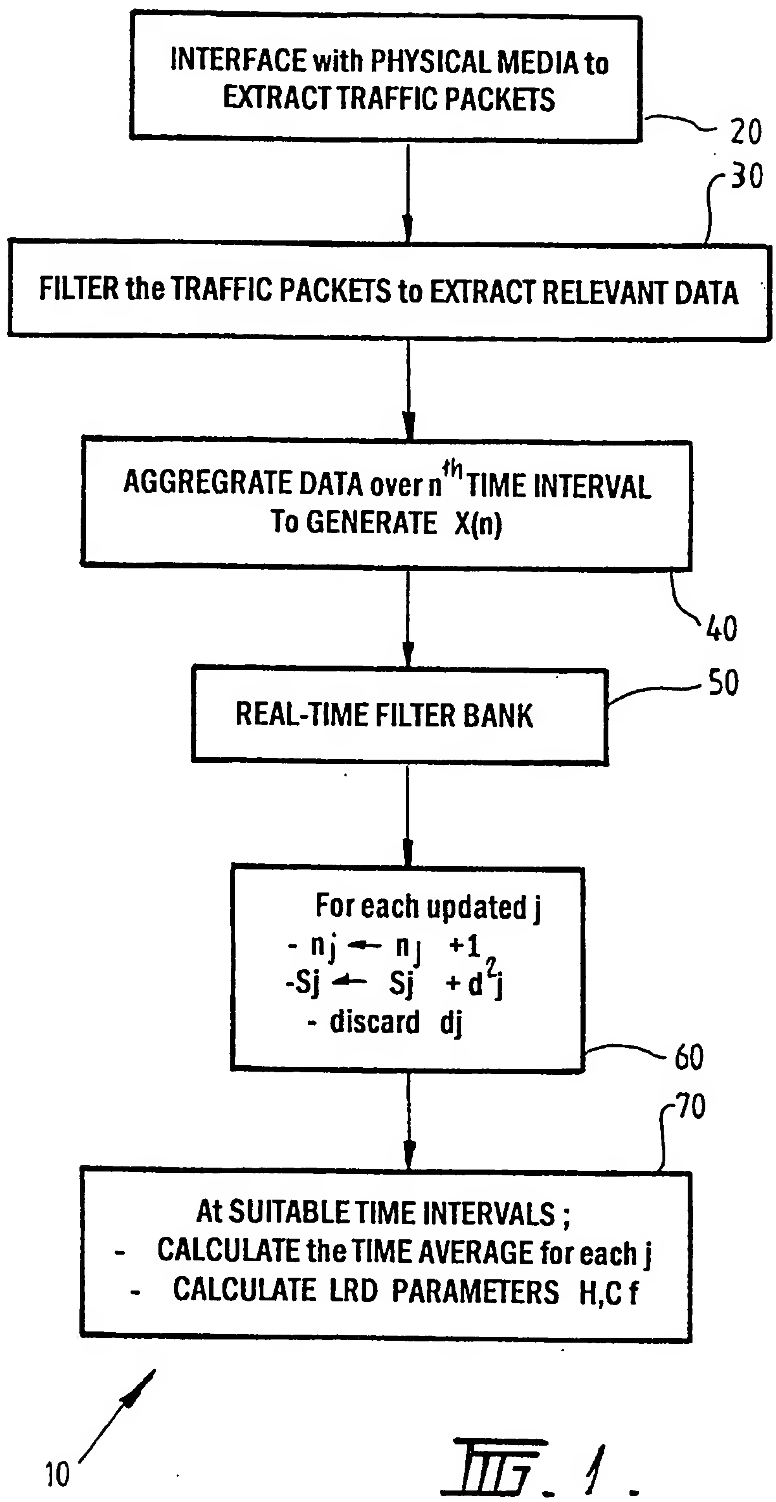

INTERFACE with PHYSICAL MEDIA to EXTRACT TRAFFIC PACKETS
20
30
FILTER the TRAFFIC PACKETS to EXTRACT RELEVANT DATA
AGGREGRATE DATA over n^th TIME INTERVAL To GENERATE X(n)
40
50
REAL-TIME FILTER BANK
For each updated j
- nj ← nj +1
-Sj ← Sj + d²j
- discard dj
60
70
At SUITABLE TIME INTERVALS ;
- CALCULATE the TIME AVERAGE for each j
- CALCULATE LRD PARAMETERS H,C f
10

FIG. 1.

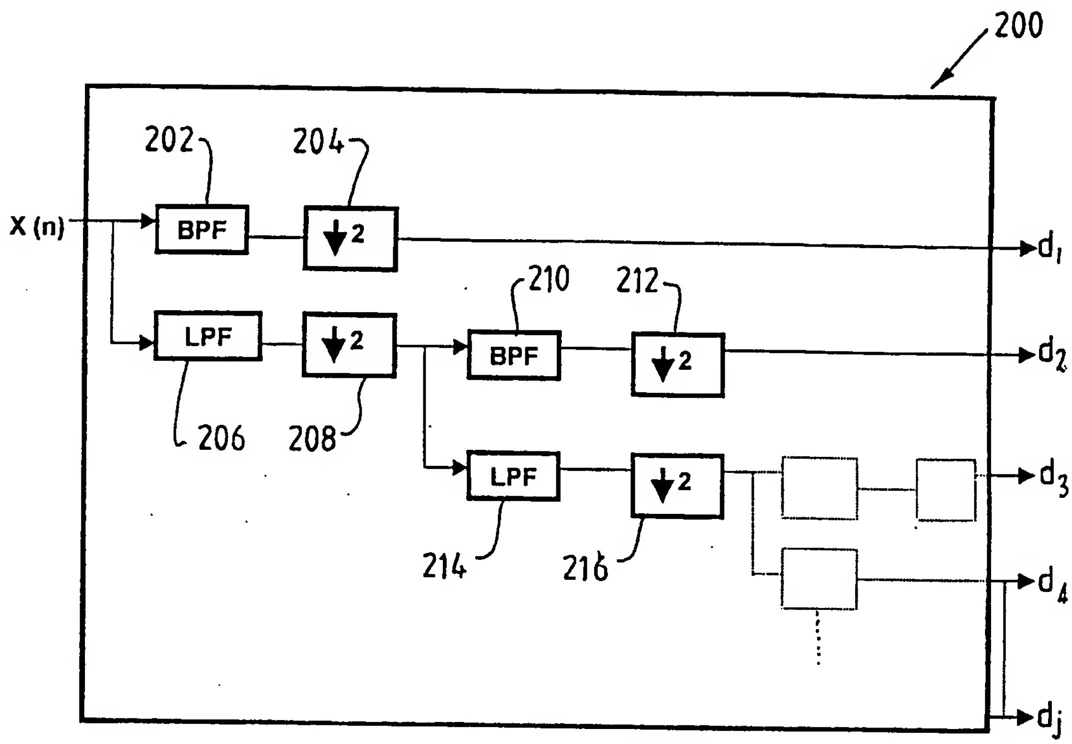

200
202
204
X (n)
BPF
↓2
d_1
210
212
LPF
↓2
BPF
↓2
d_2
206
208
LPF
↓2
d_3
214
216
d_4
d_j

FIG. 2.

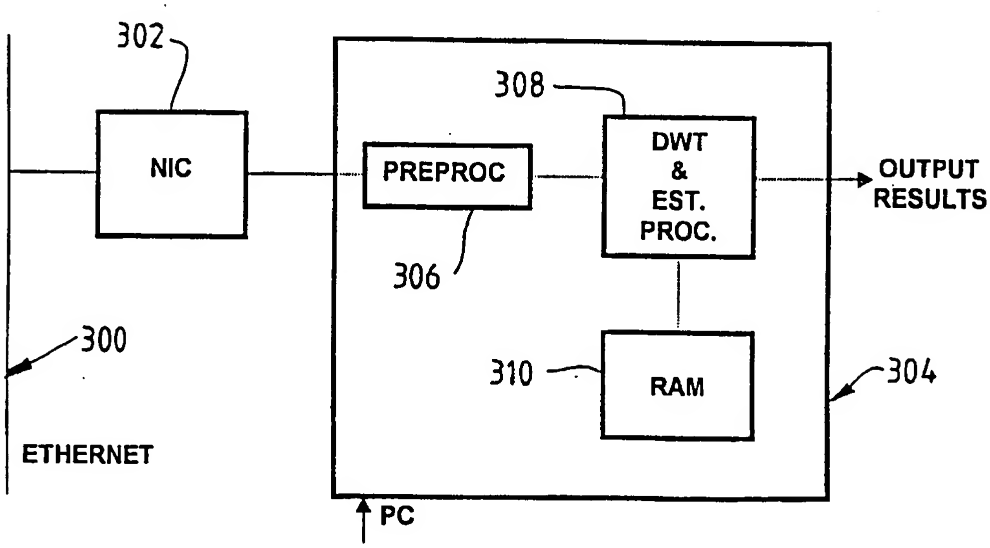

302
NIC
308
PREPROC
DWT & EST. PROC.
OUTPUT RESULTS
306
300
310
RAM
304
ETHERNET
PC

FIG. 3.

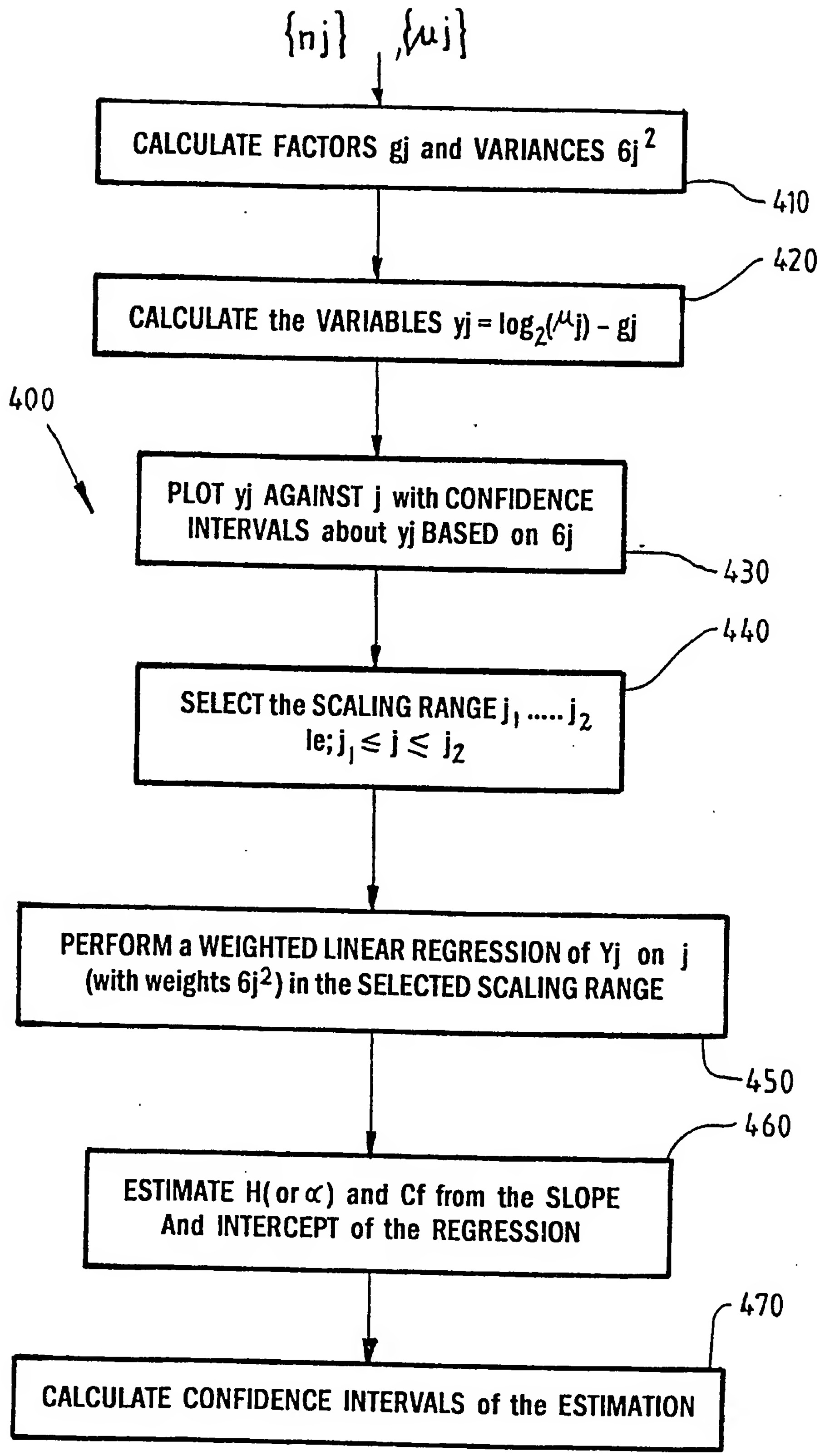

{nj} ,{μj}
CALCULATE FACTORS gj and VARIANCES 6j 2
410
420
CALCULATE the VARIABLES yj = log2(μj) - gj
400
PLOT yj AGAINST j with CONFIDENCE INTERVALS about yj BASED on 6j
430
440
SELECT the SCALING RANGE j1 ..... j2 le; j1 ≤ j ≤ j2
PERFORM a WEIGHTED LINEAR REGRESSION of Yj on j (with weights 6j2) in the SELECTED SCALING RANGE
450
460
ESTIMATE H(or α) and Cf from the SLOPE And INTERCEPT of the REGRESSION
470
CALCULATE CONFIDENCE INTERVALS of the ESTIMATION

FIG. 4.

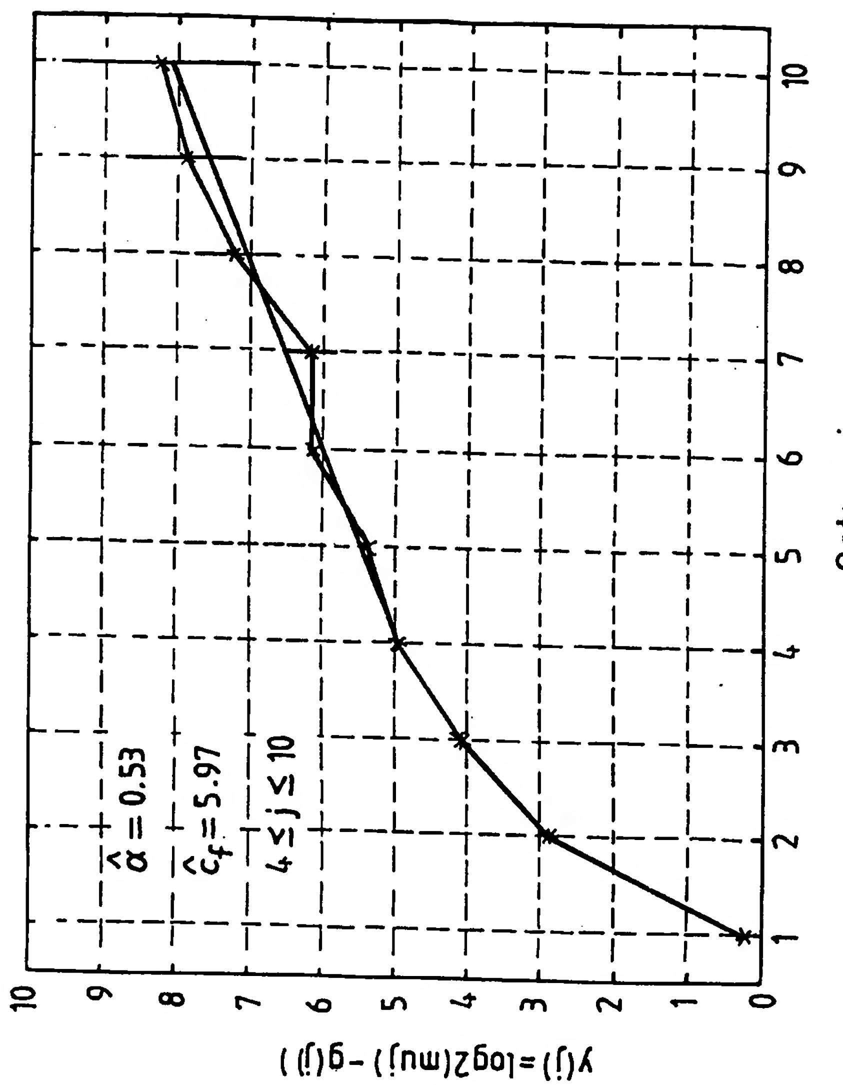

$\hat{\alpha} = 0.53$
$\hat{c}_f = 5.97$
$4 \leq j \leq 10$
Octave j
y(j) = log2(muj) - g(j)
0
1
2
3
4
5
6
7
8
9
10

FIG. 5.

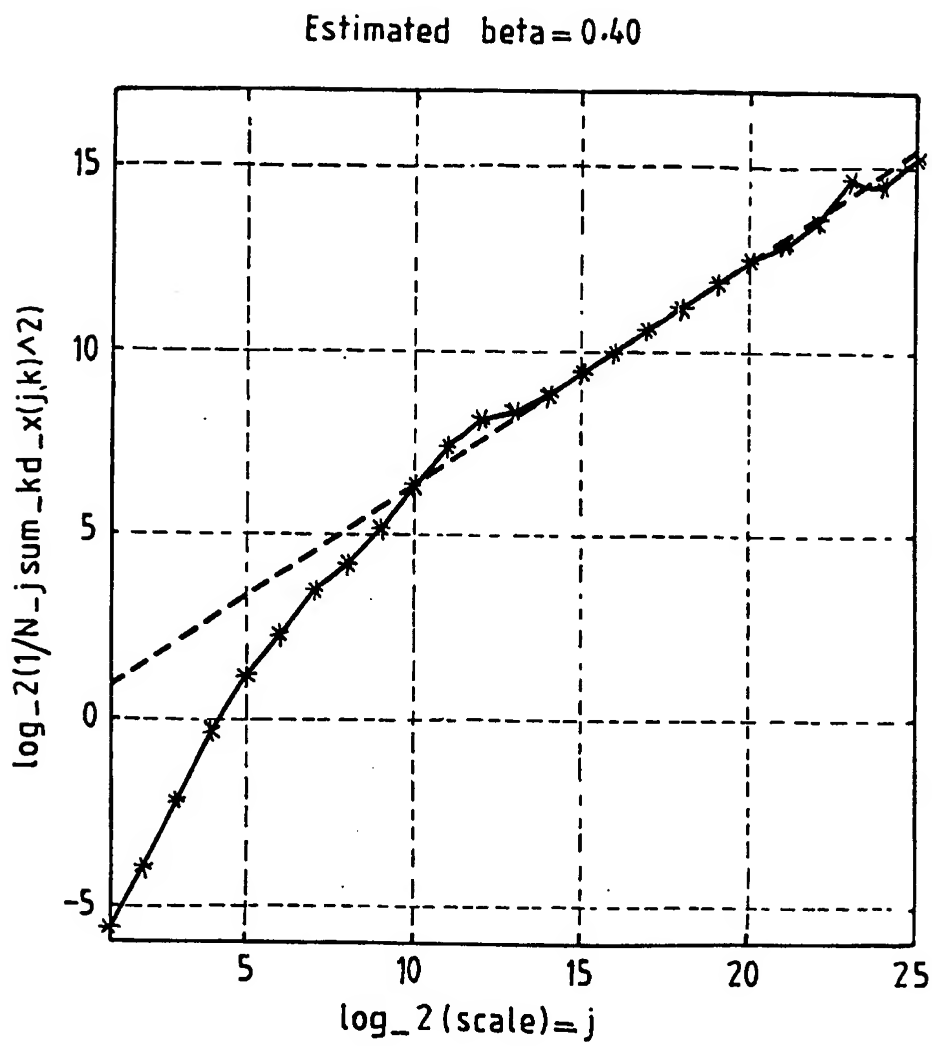

Estimated beta = 0.40
15
10
5
0
-5
log_2(1/N_j sum_kd_x(j,k)^2)
5
10
15
20
25
log_2 (scale) = j

FIG. 6.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 99/00077

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl^{6}: H04L 12/56, 12/413; G06F 17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
WHOLE IPC

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT
INSC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	IEEE Global Telecommunications Conference, Phoenix AZ, USA, 3-8 November 1997, Zhong Fan et al., "Self-Similar Traffic Generation and Parameter Estimation Using Wavelet Transform", pp 1419-1423	1-32

☐ Further documents are listed in the continuation of Box C

☐ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 March 1999	23 MARCH 1999 (23.03.99)

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929	S. AGGARWAL Telephone No.: (02) 6283 2192